

## Ders 1 : Risk Fonksiyonu ve Karar Fonksiyonu Seçimi I

*Dersi anlatan: İhsan Karabulut**Notları yazan: İhsan Karabulut*

**Not:** *LaTeX ders kalıbı UC Berkeley EECS Bölümü'nündür.*

**Uyarı:** *Bu ders notları formal yayınların tabi olduğu kanun, yönetmelik, kural ve esaslar dışındadır. Ders dışında herhangi bir şekilde kopya edilmesi, çoğaltılması, yayımlanması yalnızca bu notları hazırlayan ve yazanın iznini gerektirir.*

**Tanım.**  $X$ ,  $f(x; \theta) = P_\theta(X = x)$  olasılık fonksiyonuna sahip destek kümesinde  $x_i$  değerlerini alan bir rasgele değişken ve  $d$  verilen bir karar fonksiyonu olmak üzere

$$\begin{aligned} R(\theta, d) &= E(l(\theta, d(X))) \\ &= \sum_i l(\theta, d(x_i))f(x_i; \theta) \end{aligned}$$

beklenen değerine **risk fonksiyonu** denilir.

**Not.** Rasgele değişkenin mutlak sürekli bir rasgele değişken olması halinde yukarıdaki tanımlama

$$R(\theta, d) = \int_{-\infty}^{\infty} l(\theta, d(x))f(x; \theta)dx$$

olarak yapılacaktır.

Risk fonksiyonu kayıp fonksiyonu hem de  $X$  rasgele değişkeninin olasılık fonksiyonu ya da olasılık yoğunluk fonksiyonu nedeniyle doğanın durumu  $\theta \in \Theta$ ya bağlıdır. Bu nedenle  $R(\theta, d)$  verilen  $d$  karar fonksiyonu için  $\theta$  doğa durumundaki risk fonksiyonu değeri olarak değerlendirilir.

Rasgele gözlemlerin karar verme sürecine katılmasıyla karar verici artık bir istatistikçi olacaktır. Verilecek karar özünde rasgelelik olgusunun bulunduğu rasgele bir deneyin sonucuna göre eylem-

lerin seçilmesidir. Bunun için öncelikle bir karar kuralı-bir fonksiyon- yani sade yada karma bir karar fonksiyonu seçilecektir. **Bu seçim doğanın durumunun bilinmediği, ancak bilinen bir risk fonksiyonunun varlığı altında yapacaktır.** Karar vermenin bütün unsurları problem-leri bu durumda da varlığını sürdürür. Eğer veri uygun ve yerinde (akıllıca) kullanılırsa daha az kayıpların yaşandığı kararlar verilebilecektir, gözlemleri kullanarak karar vermenin amacı budur. Karar fonksiyonunun belirlenmesinde kullanılacak ilkeler, veri olmaksızın sade ya da karma eylem belirlemek için kullanılan minimaks ve Bayes ilkeleri olacaktır.

Karar fonksiyonunun seçimi için bütün sade ve karma karar fonksiyonlarına karşılık her  $\theta \in \Theta$  durumunda karşılık gelen  $(R(\theta_1, d), R(\theta_2, d), R(\theta_3, d), \dots, R(\theta_m, d))$  noktaların kümesi yani bir konveks kümenin elde olması gereklidir.

**Örnek.**(Asansör problemi)  $d_3$  sade karar fonksiyonu için  $\theta_1$  durumunda  $R(\theta_1, d_3)$  ve  $\theta_2$  durumunda  $R(\theta_2, d_3)$  risk fonksiyonu değerleri aşağıdaki gibi hesaplanacaktır.  $R(\theta_1, d_3)$  ve  $R(\theta_2, d_3)$ 'nin  $R(\theta, d_3)$ 'nin karar fonksiyonunun alacağı değerler olduğunu hatırlanmalıdır.

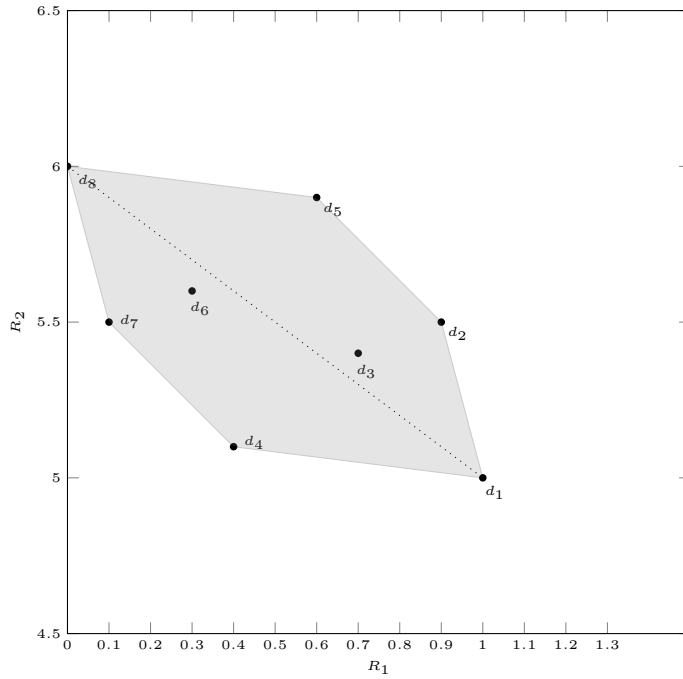
$$\begin{aligned}
 R(\theta_1, d_3) &= l(\theta_1, d_3(x_1))P_{\theta_1}(X=x_1)+l(\theta_1, d_3(x_2))P_{\theta_1}(X=x_2)+l(\theta_1, d_3(x_3))P_{\theta_1}(X=x_3) \\
 &= l(\theta_1, d_3(0))P_{\theta_1}(X=0)+l(\theta_1, d_3(1))P_{\theta_1}(X=1)+l(\theta_1, d_3(2))P_{\theta_1}(X=2) \\
 &= l(\theta_1, a_2)P_{\theta_1}(X=0)+l(\theta_1, a_1)P_{\theta_1}(X=1)+l(\theta_1, a_2)P_{\theta_1}(X=2) \\
 &= 1 \times 0.6+0 \times 0.3+1 \times 0.1 \\
 &= 0.7
 \end{aligned}$$

$$\begin{aligned}
 R(\theta_2, d_3) &= l(\theta_2, d_3(x_1))P_{\theta_2}(X=x_1)+l(\theta_2, d_3(x_2))P_{\theta_2}(X=x_2)+l(\theta_2, d_3(x_3))P_{\theta_2}(X=x_3) \\
 &= l(\theta_2, d_3(0))P_{\theta_2}(X=0)+l(\theta_2, d_3(1))P_{\theta_2}(X=1)+l(\theta_2, d_3(2))P_{\theta_2}(X=2) \\
 &= l(\theta_2, a_2)P_{\theta_2}(X=0)+l(\theta_2, a_1)P_{\theta_2}(X=1)+l(\theta_2, a_2)P_{\theta_2}(X=2) \\
 &= 5 \times 0.1+6 \times 0.4+5 \times 0.5 \\
 &= 5.4
 \end{aligned}$$

Benzer hesaplamalar diğer sade  $d_j$  karar fonksiyonları için de yapıldığında sade karar fonksiyonlarına ait risk fonksiyonu değerleri ile aşağıdaki tablo oluşturulur:

	$R(\theta_i, d_j)$							
	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$d_6$	$d_7$	$d_8$
$\theta_1$	1.0	0.9	0.7	0.4	0.6	0.3	0.1	0.0
$\theta_2$	5.0	5.5	5.4	5.1	5.9	5.6	5.5	6.0

$(1.0, 5.0), (0.9, 5.5), (0.7, 5.4), (0.4, 5.1), (0.6, 5.9), (0.3, 5.6), (0.1, 5.5), (0, 6)$  noktaları- nın  $R^2$  de oluşturduğu en küçük konveks kabuk Şekil'de yer alan grafikte gösterilmiştir. Konveks kabuk tüm sade ve karma karar fonksiyonlarına ilişkin risk fonksiyonu değerlerinin  $(R_1, R_2) = (R(\theta_1, d), R(\theta_2, d))$  -vektörlerinin- kümesidir.



Şekil 1.1: Asansör probleminde  $R(\theta, d)$  risk fonksiyonlarına ait  $(R_1, R_2)$  risk noktalarının konveks kabuğu.

Veri karar verme sürecine katılmadığı durumda bütün sade ve karma eylemlere ait beklenen  $(L_1, L_2)$  kayıpları  $(R_1, R_2)$  grafiğinde  $d_1$  ile  $d_8$  karar fonksiyonlarına ait  $(R(\theta_1, d_1), R(\theta_2, d_1))$  ile  $(R(\theta_1, d_8), R(\theta_2, d_8))$

noktaları arasındaki doğru parçası üzerinde yer alır, bu doğru parçası Sekil' deki noktalarla işaretlenmiştir.  $d_4, d_6$  ve  $d_7$  sade karar fonksiyonlarının risklerinin  $(R(\theta_1, d_1), R(\theta_2, d_1))$  ile  $(R(\theta_1, d_8), R(\theta_2, d_8))$  noktaları arasındaki doğru parçası üzerinde yer alan karar fonksiyonlarına baskın olduğu görülecektir; verinin kullanılmasıyla verinin kullanılmadığı zaman karar verici için kabul edilebilir olan sade ya da eylemlere baskın riski düşük karar fonksiyonları elde edilmiştir ( $a_1$  ve  $a_2$  sade eylemlerine karşılık gelen ve kabul edilebilir olan  $d_1$  ve  $d_8$  dışında) bu verinin kullanılması ile sağlanmıştır. Diğer taraftan yine verini kullandığı  $d_2, d_3, d_5$  sade karar fonksiyonları ile doğru parçasının daha üzerinde olan noktaların temsil ettiği karar fonksiyonlarının riskleri veri kullanılıyor olmasına karşın daha büyüktür ve kabul edilebilir de değildirler.

Yukarıdaki gözlemler tipik olup doğanın durumu ile ilgili doğanın durumuyla ilgili rasgele gözlemlerin akıllıca kullanımı ile risklerin sıfır noktasına kadar çekilebileceği sonucu çıkarılır.

Veri doğanın durumu ile ilgili değilse - **doğanın durumuna ilişkin informas- yona (malumata) sahip değilse** -  $(R_1, R_2)$  noktaları sadece  $(L_1, L_2)$  noktalarının konveks olacaktlardır. Örneğin, rasgele değişken  $X$  in dağılımı doğanın durumu  $\theta \in \Theta$  dan bağımsız ise

$$P_{\theta}(X = x) = f(x; \theta) = h(x)$$

İki doğa durumunun var olduğu bir karar verme probleminde verilen karar fonksiyonu  $d$  için riskler  $i = 1, 2$  olmak üzere

$$R(\theta_i, d) = E(l(\theta_i, d(X))) = \sum_j h(x_j)l(\theta_i, d(x_j))$$

olacaktır.  $d(x_j), a_i \in \mathcal{A}$  olacak bu nedenle  $l(\theta_1, a_i) = L_1$  ve  $l(\theta_2, a_i) = L_2$  olarak gösterilebilecektir.  $(R_1, R_2)$ 'nin bileşenleri bunların birer konveks kombinasyonu olacak ve gözlem yapıp karar karar vermekle değişen bir şey olmayacak; beklenen kayıpların kümesi risk noktalarının kümesine taşınmış olmayacaktır.

## 1.1 Karar Fonksiyonu Seçimi

Baskın olma, denklik ve kabul edilebilirlik tanımları beklenen kayıplar için nasıl tanımlanıyorsa riskler için de benzer olarak tanımlanır.

$d$ ,  $d^*$  karar fonksiyonları kümesinde yer alan herhangi iki karar fonksiyonu olsun.  $R(\theta_1, d) \leq R(\theta_1, d^*)$ ,  $R(\theta_2, d) \leq R(\theta_2, d^*)$ ,  $R(\theta_3, d) \leq R(\theta_3, d^*)$ ,  $\dots$ ,  $R(\theta_m, d) \leq R(\theta_m, d^*)$  ise  $d$  **karar fonksiyonu en az  $d^*$  karar fonksiyonu kadar iyidir** denilir.

$R(\theta_1, d) = R(\theta_1, d^*)$ ,  $R(\theta_2, d) = R(\theta_2, d^*)$ ,  $R(\theta_3, d) = R(\theta_3, d^*)$ ,  $\dots$ ,  $R(\theta_m, d) = R(\theta_m, d^*)$  ise  $d$  **karar fonksiyonu  $d^*$  karar fonksiyonuna denktir** denilir.

İki karar fonksiyonu denk ise bu karar fonksiyonların tanımları farklı olsalar da,  $R^k$  de aynı nokta ile temsil edilebilir. Örneğin bir sade karar fonksiyonu bir karma karar fonksiyonuna denk olabilir ve  $R^k$ 'de aynı nokta ile temsil edilebilirler.

$d$  en az  $d^*$  karar fonksiyonu kadar iyi ve  $d^*$  karar fonksiyonuna denk değilse  $d$  **karar fonksiyonu  $d^*$  karar fonksiyonuna baskındır** denilir.  $d$  karar fonksiyonuna baskın bir karar fonksiyonu yoksa  $d$  **kabul edilebilir** bir karar fonksiyonu denilir.

Bu kez problem kayıp fonksiyonunun bilinmesi ve doğanın hangi durumda olduğunu bilmeksizin bir (veya daha fazla ) eylemin seçimine benzer olarak, risk fonksiyonlarının bilinmesi fakat doğanın hangi durumda olduğunu bilmeden karar fonksiyonları kümesinden **bir (veya daha fazla) karar fonksiyonunun** seçimidir. Önceden de ifade edildiği gibi seçim yapmak için yine minimaks ve Bayes ilkeleri kullanılacaktır.

**Minimaks karar fonksiyonu**  $d$ 'nin belirlenmesi için, önce verilen her  $d$  karar fonksiyonunun hangi  $\theta_i \in \Theta$  doğa durumunda  $R(\theta, d)$  risk fonksiyonunun en büyük (maksimum) değerli olduğu belirlenir, bu işlem

$$M(d) = \max_{\theta \in \Theta} R(\theta, d)$$

ile gösterilsin.  $M(d)$  ler arasında en küçüğü (minimumu) gerçekleyen  $d$  karar fonksiyonu

$$\min_{\theta \in \Theta} \max R(\theta, d)$$

minimaks karar fonksiyonu olarak belirlenir. İki doğa durumunun varlığında yine grafiksel minimaks çözümlenmesi kullanılabilir, ancak doğa durumlarının sayısı daha büyük olursa genel olarak çözüme ulaşmak için sayısal çözümlenmeye gereksinim duyulur.

**Bayes karar fonksiyonu** da  $\theta \in \Theta$  doğa durumları için verilen bir  $g(\theta)$  önsel dağılım altında belirlenebilir. Verilen bir  $d$  karar fonksiyonu için  $\theta$  nin bir fonksiyonu olarak risk fonksiyonunun önsel dağılıma göre beklenen değeri doğa durumu sayılabilir olduğunda

$$B(d) = E(R(\theta, d)) = \sum_j R(\theta_j, d)g(\theta_j)$$

olacaktır. Benzer olarak doğa durumları  $(-\infty, \infty)$  reel sayılar kümesinde veya reel sayılar kümesinde bir aralıkta değer alıyorsa ve  $g(\theta)$  önsel dağılımı mutlak sürekli bir r.d. için olasılık yoğunluk fonksiyonu ise risk fonksiyonunun önsel dağılıma göre beklenen değeri

$$B(d) = E(R(\theta, d)) = \int_{-\infty}^{\infty} R(\theta, d)g(\theta)d\theta$$

olacaktır.  $B(d)$  **Bayes riski** olarak adlandırılacaktır. Bayes riskini minimum yapan karar fonksiyonu

$$B(d^*) = \min_d B(d)$$

$d^*$  **Bayes karar fonksiyonu** olarak adlandırılacaktır.

## Ders 2 : Risk Fonksiyonu ve Karar Fonksiyonu Seçimi II

*Dersi anlatan: İhsan Karabulut**Notları yazan: İhsan Karabulut*

Bu tanımlamalar kayıp fonksiyonunun kullanılması yerine pişmanlık fonksiyonları kullanılarak yapılabilir. Önceden de değinildiği gibi bu değişiklik minimaks ilkesinin uygulanması durumunda karara konulacak çözümde bir farklılık oluşturabilecek ancak Bayes yöntemi kullanıldığında bir farklılık oluşturmayacaktır. Hatırlanacak olursa pişmanlık fonksiyonu

$$r(\theta, a_i) = l(\theta, a_i) - \min_a l(\theta, a)$$

dır. Tanımda yer alan  $a_i$  yerine herhangi bir  $d(x)$  karar fonksiyonu konulduğunda karar fonksiyonları için pişmanlık fonksiyonu

$$r(\theta, d(x)) = l(\theta, d(x)) - \min_a l(\theta, a)$$

olmalıdır. Dikkat edilirse bu tanımlamada  $\min_a l(\theta, a)$  teriminde bir değişiklik yoktur. Çünkü, hangi  $d$  karar fonksiyonu olursa olsun karar fonksiyonunun belirlediği bir  $a$  eylemi olacak ve doğacak pişmanlık yine  $\min_a l(\theta, a)$  teriminden oluşan farkla belirlenmiş olacaktır. Pişmanlık riski ise

$$\begin{aligned} E(r(\theta, d(X))) &= E(l(\theta, d(X))) - \min_a l(\theta, a) \\ &= R(\theta, d) - \min_a l(\theta, a) \end{aligned}$$

olur. Herhangi bir karar fonksiyonu  $d$  ve  $X = x$  için  $l(\theta, d(x)) \geq \min_a l(\theta, a)$  olduğundan  $E(l(\theta, d(X))) \geq \min_a l(\theta, a)$  olacaktır bu nedenle

$$\min_d E(l(\theta, d(X))) = \min_d R(\theta, d(X)) \geq \min_a l(\theta, a)$$

olmalıdır.

Diğer taraftan bir  $a^*$  eylemi  $l(\theta, a)$  kaybını en küçük yapar ve karar verici  $d^*(x) = a^*$  olan bir  $d^*$  karar fonksiyonu kullanıyorsa bu kez

$$\min_d R(\theta, d) \leq R(\theta, d^*) = E(l(\theta, a^*)) = l(\theta, a^*) = \min_a l(\theta, a)$$

olacaktır. Bu da yukarıdaki tanıma ilişkin **pişmanlık riskinin**

$$E(r(\theta, d(X))) = R(\theta, d) - \min_d R(\theta, d)$$

olarak yeniden yazabileceği anlamına gelir.

**Örnek. (Asansör problemi)** Sade karar fonksiyonları arasından Bayes ve minimaks karar fonksiyonları **pişmanlık riski kullanılarak** elde edilecektir. Sade karar fonksiyonları  $d_1, d_2, \dots, d_8$  olmak üzere

$$\begin{aligned} \min_d R(\theta_1, d) &= R(\theta_1, d_8) = 0 \\ \min_d R(\theta_2, d) &= R(\theta_2, d_1) = 5 \end{aligned}$$

olduğundan sade karar fonksiyonlarının risklerinden tanımlanan pişmanlık riskleri  $E(r(\theta, d(X))) = R(\theta, d) - \min_d R(\theta, d)$  her iki doğa durumunda aşağıdaki tabloda sunulmuştur.

	$E(r(\theta_i, d_j(X)))$							
	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$d_6$	$d_7$	$d_8$
$\theta_1$	1.0	0.9	0.7	0.4	0.6	0.3	0.1	0.0
$\theta_2$	0.0	0.5	0.4	0.1	0.9	0.6	0.5	1.0

Aşağıda sade karar fonksiyonları arasından minimaks karar fonksiyonu elde edilecektir. Bunun için  $M(d) = \max_{\theta} R(\theta, d)$  yerine pişmanlık riski kullanılarak, verilen her sade karar fonksiyonu için en büyük pişmanlığın hangi doğa durumunda olduğu

$$M(d) = \max_{\theta} E(r(\theta, d(X))) = \max_{\theta} \{R(\theta, d) - \min_d R(\theta, d)\}$$



ile saptanmalıdır. Yukarıdaki son tabloda verilenleri kullanarak  $d_1$  karar fonksiyonu için

$$\begin{aligned}
 M(d_1) &= \max_{\theta} \{E(r(\theta, d_1(X)))\} \\
 &= \max\{E(r(\theta_1, d_1(X))), E(r(\theta_2, d_1(X)))\} \\
 &= \max\{1, 0\} \\
 &= 1
 \end{aligned}$$

bulunacaktır. Diğer sade karar fonksiyonları için bulunan  $M(d_j)$  değerleri aşağıdaki tablonun ilk satırında verilmiştir. Bunlar arasından -karar fonksiyonları üzerinden- minimum bulunarak minimaks karar fonksiyonu ya da karar fonksiyonları bulunur:

$$\begin{aligned}
 \min_d \{M(d)\} &= \min\{M(d_1), M(d_2), \dots, M(d_8)\} \\
 &= M(d_4) \\
 &= 0.4
 \end{aligned}$$

olup  $d_4$  sade karar fonksiyonları arasından minimaks olan karar fonksiyonudur.

Sade karar fonksiyonları arasından Bayes kararı fonksiyonunu belirlemek için doğa durumları için önsel dağılım  $g(\theta_1) = w_1 = 0.8$ ,  $g(\theta_2) = w_2 = 0.2$  olsun.  $d_2$  sade karar fonksiyonunun Bayes

pişmanlık riski  $B(d_2) = E(r(\theta, d_2))$  aşağıda elde edilmiştir:

$$\begin{aligned}
 B(d_2) &= \sum_{i=1}^2 E(r(\theta_i, d_2(X)))g(\theta_i) \\
 &= \sum_{i=1}^2 (R(\theta_i, d_2) - \min_d R(\theta_i, d))g(\theta_i) \\
 &= (R(\theta_1, d_2) - 0) \times 0.8 + (R(\theta_2, d_2) - 5) \times 0.2 \\
 &= 0.9 \times 0.8 + 0.5 \times 0.2 \\
 &= 0.82
 \end{aligned}$$

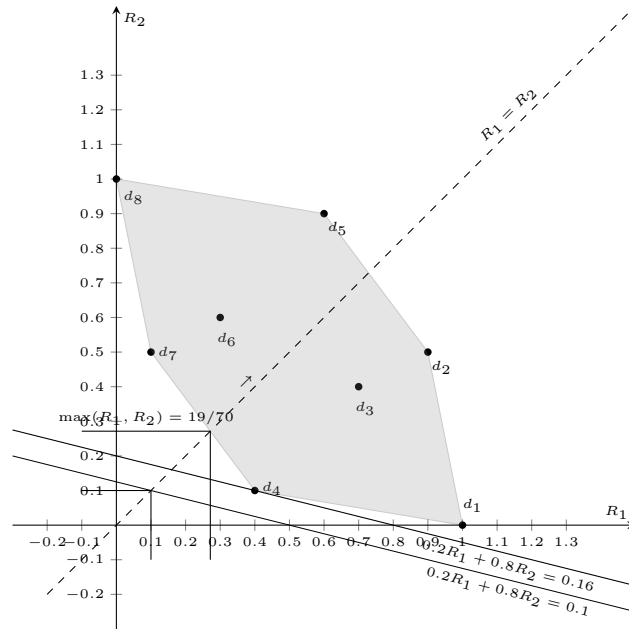
Diğer sade karar fonksiyonları için de benzer olarak  $B(d_j)$  Bayes pişmanlık risk değerleri bulunur. Bunlar da yine aşağıdaki tablonun ikinci satırında verilmiştir. Buna göre sade karar fonksiyonları arasından  $d_7$  sade karar fonksiyonu Bayes pişmanlık riskini en küçük yapar. Sade karar fonksiyonları arasında  $d_7$  Bayes karar fonksiyonudur ve Bayes pişmanlık riski 0.18 dir.

$$\min\{B(d_1), B(d_2), \dots, B(d_8)\} = B(d_7) = 0.18$$

	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$d_6$	$d_7$	$d_8$
$M(d)$	1.00	0.90	0.70	0.40	0.90	0.60	0.50	1.00
$B(d)$	0.80	0.82	0.64	0.34	0.66	0.36	0.18	0.20

**Tablo.** Asansör probleminde pişmanlık riskleri kullanılarak sade karar fonksiyonları arasından minimaks ve önsel dağılım (0.8, 0.2) verildiğinde Bayes karar fonksiyonlarının bulunması.

Doğa durumları uzayında iki doğa durumunun olması halinde geometrik yöntemi kullanılarak bütün sade ve karma karar fonksiyonları içinden minimaks ve Bayes karar fonksiyonları belirlenebilir. Asansör örneğinde de pişmanlık risklerine ait konveks kabuk Şekil'deki gibi elde edilir. Örnek üzerinde önce minimaks karar fonksiyonunu belirleyelim. Şekil üzerinde de görüleceği gibi minimaks karar fonksiyonu  $d_4$  ve  $d_7$  sade karar fonksiyonlarının bir karması olan karar fonksiyonu olacaktır.  $d_4$  karar fonksiyonunun  $p$  olasılıkla uygulanmasına veya  $d_7$  karar fonksiyonunun  $1 - p$  olasılıkla uygulanmasına karar verildiğinde karma karar fonksiyonuna ilişkin pişmanlık riski beklenen değeri-



Şekil 2.2: Asansör probleminde tüm sade ve karma karar fonksiyonları arasındaki minimaks ve Bayes karar fonksiyonlarının grafik ile belirlenmesi. Bayes ilkesine göre çözüm için önsel dağılım  $g(\theta_0) = 0.20$  ve  $g(\theta_1) = 0.80$  dir.

vektörü-  $x = y$  olacak şekilde

$$\begin{pmatrix} x \\ y \end{pmatrix} = p \begin{pmatrix} 0.4 \\ 0.1 \end{pmatrix} + (1-p) \begin{pmatrix} 0.1 \\ 0.5 \end{pmatrix}$$

olarak ifade edilebilecektir.

$$0.4p + 0.1(1-p) = 0.1p + 0.5(1-p)$$

$p \in [0, 1]$  ye göre çözüldüğünde  $p = 4/7$  bulunur. Bu değer  $x = 0.4p + 0.1(1-p)$  veya  $y = 0.1p + 0.5(1-p)$  eşitliklerinden birinde yerine konulduğunda söz konusu minimaks karma eyleme ilişkin beklenen pişmanlık riskinin  $19/70$  olduğu bulunur. Bu probleme ilişkin veri olmaksızın yapılan çözümlemede minimaks kaybı  $1/2$  bulunur. Biri risk diğeri beklenen kayıp olmakla birlikte veri kullanılarak  $(1/2) - (19/70) = 16/70$  kadar bir "kayıp" azaltılmıştır yorumunu yapılabilir.

## Kaynaklar

- [1] D. BLACKWELL and M.A. GIRSHICK (1979), Theory of games and statistical decisions, *Dover Publications*, New York.
- [2] H. CHERNOFF and L. E. MOSES (1986), Elementary Decision Theory, *Dover Publications*, New York.
- [3] B. W. LINDGREN (1971), Elements of Decision Theory, *Macmillan Company*, New York.
- [4] A. SABUNCUOĞLU (2004?), Analitik Geometri, 2. Baskı, *Nobel Yayınları* , Ankara?